DOCUMENT RESUME

ED 432 183 HE 032 176

AUTHOR Azevedo, Roger
TITLE Expertise in Radiology: Accounting for the Evidence and Implications for Instruction.
PUB DATE 1999-04-00
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Ontario, Canada, April 19-23, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clinical Diagnosis; Cognitive Processes; Computer Assisted Instruction; Error Patterns; Graduate Medical Education; Higher Education; Physicians; *Problem Solving; *Professional Education; *Radiology; *Specialists
IDENTIFIERS *Expertise; Mammography

ABSTRACT

The problem-solving strategies used by 10 radiologists and 10 radiology residents during the interpretation of difficult mammograms were examined under two experimental conditions. In the "authentic" condition, standard unmarked mammograms were used. On a second set of the same cases for the "augmented" condition, mammographic findings were highlighted. Verbal protocols revealed that mammography interpretation predominantly used data-driven or mixed-strategies but depended on case typicality and clinical experience. Radiologists scanned the cases significantly faster than residents, but no group differences were found in the number of findings, observations, and diagnoses. The groups (1) used the same types of operators, control processes, and diagnostic plans; (2) committed the same number of errors; and (3) both groups committed case-dependent errors. Overall, the results indicated that mammogram interpretation is a well-constrained visual cognitive task. Results were applied to the design of a computer-based tutor for mammogram interpretation. (Author, DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 432 183

Running Head: PROBLEM SOLVING IN RADIOLOGY

Expertise in Radiology:
Accounting for the Evidence and Implications for Instruction

Roger Azevedo

Carnegie Mellon University

Department of Psychology

Baker Hall, #365

Pittsburgh, PA 15213

Tel: (412) 268-4239

Fax: (412) 268-2798

E-mail: azevedo@andrew.cmu.edu

Please address all correspondence to (after August 1, 1999):

Roger Azevedo

University of Maryland

Department of Human Development

3304 Benjamin Building

College Park, MD 20742-1131

Tel: (301) 405-2827

Fax: (301) 405-2891

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Roger Azevedo

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

Paper presented at the annual conference of the [1] American Educational Research Association, Montréal, Quebéc, Canada (April 19-23, 1999)

HE 032 176

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

Abstract

The problem solving strategies used by 10 radiologists and 10 radiology residents during the interpretation of difficult mammograms were examined under two experimental conditions. In the *authentic* condition, standard unmarked mammograms were used. Mammographic findings were highlighted on a second set of the same cases for the *augmented* condition. Verbal protocols revealed that mammography interpretation predominantly used data-driven or mixed-strategies but depended on case typicality and clinical experience. Radiologists scanned the cases significantly faster than residents but no group differences were found in the number of findings, observations, and diagnoses. The groups (1) used the same types of operators, control processes, and diagnostic plans, (2) committed the same number of errors, and (3) both committed case-dependent errors. Overall, the results indicated that mammogram interpretation is a well-constrained visual cognitive task. The results were applied to the design of a computer-based tutor for mammogram interpretation. Future empirical directions include building a more comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation. The results of this study contribute to the fields of educational psychology and applied cognitive science.

Problem Solving in Radiology:
Accounting for the Evidence and Implications for Instruction

Introduction

Breast cancer is the leading form of cancer diagnosed in North American women (excluding non-melanoma skin cancer), accounting for about 30% of all new cases (Gaudette, Silberberger, Altmayer & Gao, 1996). After the age 30, incidence rates begin to rise, and the highest rates are among women aged 60 and over. Incidence rates have increased slowly and steadily since 1969, rising most rapidly among women aged 50 and over. Mammographic screening has become an accepted means of substantially reducing breast cancer mortality. However, 11% to 25% of cancers are overlooked by radiologists on initial screening mammograms (Goergen, Evans, Cohen & MacMillan, 1997). The high incidence rates together with the rate of misdiagnoses make this an alarming problem which is associated with societal, ethical, and additional medical concerns. Given the scope and seriousness of the problem, it is evident that any promising means for alleviating it should be investigated. Aside from the health and medical sciences, other disciplines such as cognitive science can contribute to an understanding of this domain. The cognitive components involved in proficient mammogram interpretation can be modeled and the results can subsequently be used to improve future radiological training.

The present study investigated the problem solving strategies used by staff radiologists and radiology residents during the process of diagnosing difficult breast diseases depicted on mammograms. The results of this study address existing problems associated with radiology residency training and are used as the basis for the design of the RadTutor, a computer-based learning environment to train radiology professionals in the interpretation of mammograms. Cognitive science is an interdisciplinary field that is used to build an understanding of "thinking." A basic assumption is that the mind is a computational system that constructs, manipulates, and represents symbols (Newell & Simon, 1972; Simon, 1979). Contributing disciplines such as cognitive psychology provide cognitive science with different ways of investigating the nature of "thinking," including numerous epistemological frameworks, research methodologies, and

analytical approaches (Chi, Glaser, & Farr, 1988; Ericsson, 1996; Ericsson & Simon, 1993; Ericsson & Smith, 1991; Feltovich, Ford, & Hoffman, 1997; Posner, 1989; Sternberg & Smith, 1988). Further, information processing theories, research on expert-novice differences, and the widespread use of cognitive task analysis have contributed to our understanding of learning and instruction in various educational (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995) and professional domains (e.g., Lajoie & Lesgold, 1992). As such, cognitive science offers a foundation for the study of mammogram interpretation and the results of this cognitive study can be applied to the improvement of medical training. This study investigates the underlying nature of radiological expertise in both staff radiologists (M.D.s with extensive post-residency training) and radiology residents (M.D.s completing their residency training) by focusing on their problem solving strategies. The results provide a better understanding and a performance model of mammogram interpretation which will lead to identifying better training methods.

Radiological diagnosis is complex, involving several years of acquiring formalized medical knowledge as well as many years of clinical experience. The ability to diagnose accurately necessitates the integration of several bodies of knowledge with separate organizing principles, including physiology, anatomy, pathophysiology, and projective geometry of radiography. Various theoretical frameworks postulate that the attainment of accurate visual diagnostic reasoning abilities involves the interaction between cognitive and perceptual factors. In order to adequately understand the diagnostic process, a more detailed investigation is required.

Numerous researchers, employing disparate theoretical and empirical paradigms, have investigated radiological expertise. Three basic "paradigms" that have been investigated widely include search studies, signal-detection studies, and cognitive research. Relatively few cognitive studies (Faremo, 1997; Lesgold, Feltovich, Glaser, & Wang, 1981; Lesgold, Rubinson, Feltovich, Glaser, Klopfer, & Wang, 1988; Rogers, 1992) have investigated the underlying cognitive and perceptual factors. These studies have focused mainly on the interpretation of chest x-rays. As a result, a fundamental understanding of the constitution and acquisition of expertise in other radiological sub-specialties, such as mammography interpretation, has yet to be achieved.

This will require the use of appropriate cognitive science methodologies such as protocol analysis and the study of the problem solving strategies of radiology professionals with varying levels of expertise.

Objectives of the Study

The present study investigated the problem solving strategies used by staff radiologists and radiology residents during the process of diagnosing difficult breast disease cases depicted on mammograms. The specific research objectives addressed in this study included:

(1) To construct a model of problem solving in mammogram interpretation;

(2) To determine whether staff radiologists and radiology residents use different problem solving strategies, operators, and control processes during mammogram interpretation;

(3) To determine the effects of the authentic and augmented experimental conditions for experts' and novices' on several aspects of the groups' performance; and,

(4) To determine the effects of the authentic and augmented experimental conditions on the frequency and types of errors committed by both groups. The results of this study provided an initial characterization of the cognitive processes underlying mammogram interpretation, and served as the empirical basis for the design of the RadTutor, a computerized tutor designed to train radiology residents to interpret mammograms (Azevedo & Lajoie, 1998).

Theoretical Framework

This study is based on the information processing (IP) theory typically used in expertise studies (Anderson, 1993; Newell & Simon, 1972; Posner, 1989). The basic assumption behind this framework is that the mind is a computational system that constructs, manipulates, and represents symbols. In general, participants with varying levels of expertise are asked to verbalize their "thinking processes" as they attempt to carry-out a specific task (e.g., diagnose a breast disease case). The verbal protocols are then segmented and analyzed to extract the underlying cognitive components of expertise, such as the types of knowledge (e.g., declarative), problem solving operators (e.g., hypothesis generation), control processes (e.g., diagnostic planning), problem solving strategies (e.g., data-driven), and other performance measures (e.g., error types)

used during task performance. The assumption underlying this approach is that the novice-expert progression in any domain is characterized by both quantitative and qualitative changes in problem solving strategies and knowledge representation and use. These differences can therefore be used to provide an understanding of the novice-expert differences in a domain (e.g., Lajoie & Lesgold, 1992) and also therefore have numerous pedagogical implications (e.g., Anderson, Corbett, Koedinger, & Pelletier, 1995). In accordance with the IP framework, this study investigated the underlying nature of radiological expertise in both staff radiologists (M.D.s with extensive post-residency training) and radiology residents (M.D.s completing their residency training) by focusing on their problem solving strategies and several other performance measures.

Method

Participants.

A total of 20 participants, 10 staff radiologists and 10 radiology residents from several large metropolitan university teaching hospitals participated in this study. The 10 radiologists had MD degrees and Board Certification in radiology and were affiliated with one of the teaching hospitals. They had an average of 20 years of post-residency training, an average of 14 years of mammography training, and estimated to have analyzed an average of 30,000 mammograms over the course of their medical training. The 10 residents had MD degrees and were on rotation at one of the teaching hospitals. This group comprised 3 fourth-year, and 7 fifth-year residents. All of the residents had completed 1 mammography rotation. They reported to have an average of 6 months of mammography training.

Breast Disease Cases.

Ten breast disease cases were used in this study. Cases were selected by the consulting radiologist from her teaching files. Each case was comprised of a brief clinical history and at least 4 mammograms (including the craniocaudal and mediolateral views of the left and right breasts). The cases included 3 benign and 7 malignant diseases, and these diagnoses have been confirmed by pathology reports. The cases included ones that are typically encountered in mammography textbooks and clinical research articles, atypical ones infrequently encountered in daily practice,

and ones with typical mammographic manifestations encountered in daily practice. These cases included mammographic features that ranged from fairly obvious to detect to cases that require the use of a magnifying glass to detect subtle mammographic features.

Coding Scheme.

A coding scheme was constructed based on the content analysis of breast disease and mammography, theoretical and methodological articles in cognitive science and the results of previous studies in various relevant areas such as medical cognition (Hassebrock & Prietula, 1992; Patel, Arocha, & Kaufmann, 1994; Patel & Groen, 1986), discourse processing (Bracewell & Breuleux, 1993; Breuleux, 1991; Frederiksen, 1975), and syntactical analysis (Winograd, 1983). Fifty of the 200 protocols collected were used to refine an initial coding scheme into a more comprehensive one consisting of three major categories. The major categories included *knowledge states*, *problem solving operators*, and *control processes* (Anderson, 1993; Newell & Simon, 1972). *Knowledge states* in this domain were coded as radiological observations[1], radiological findings[2], and diagnoses. *Problem solving operators* were clustered around 11 classes (e.g., hypothesis generation) and comprised a total of 30 operators. *Control processes* were comprised of diagnostic planning, goal verbalizations, and meta-reasoning.

Inter-rater reliability was established by recruiting a graduate student with experience in the area of breast disease and mammography and training her to use the coding scheme. She was instructed to independently code 40 randomly selected protocols thus yielding a reliability coefficient of .91.

Research Design.

A mixed factorial design was employed. Two levels of radiological expertise (radiologists and residents) constituted the between-subjects factor. The experimental conditions (authentic and augmented) constituted the within-subjects factor.

---

[1] units of information that are recognized as potentially relevant in the problem solving context, but do not constitute clinically useful facts

[2] composed of critical cues with particular clinical significance

Procedure.

Participants were tested individually. The experimenter provided the participant with a 1-page handout of instructions for the diagnostic task. For each case, the experimental procedure involved having the participant: (1) read the clinical history, (2) display the mammogram set on a view-box, (3) point to the mammographic findings and/or observations, (4) provide a diagnosis (or a set of differential diagnoses), and (5) discuss subsequent investigations (if necessary). The participant was instructed to "think out loud" (Ericsson & Simon, 1993) throughout the entire diagnostic process. The experimental procedure was repeated for each subject until he/she diagnosed all ten cases under the two experimental conditions (five authentic and five augmented). No time constraints were imposed.

Results

The results discussed in this section are all based on analyses of 50 randomly selected verbal protocols.

Problem Solving Model of Mammogram Interpretation.

A problem solving model of mammogram interpretation was constructed from the verbal protocol analyses. Decomposition of the complex task of mammogram interpretation resulted in a model consisting of seven steps, including: (a) reading a clinical history, (b) placing a set of mammograms on a viewbox and identifying individual mammograms in the set, (c) visually inspecting each of the mammograms either with or without the use of a magnifying glass, (d) identifying mammographic findings and observations, (e) characterizing mammographic findings and observations, (f) providing a definitive diagnosis or a set of differential diagnoses, and (g) specifying subsequent examinations (if required).

The model allows for a *"linear approach"* (e.g., from reading the clinical history to specifying subsequent examinations) and/or an *"iterative approach"* in which the results of a step may feed back to previous steps in the model. The linear approach is analogous to the use of a data-driven problem solving strategy whereby a subject reads the clinical history, scans the set of mammograms, provides a diagnosis, and specifies a subsequent examination. The iterative

approach is analogous to the use of a mixed-problem solving strategy (i.e., includes both data-driven and goal-driven problem solving strategies). For example, following the initial scanning and characterization of the mammographic findings the participant may postulate a set of differential diagnoses which will lead him/her to inspect particular area(s) of a mammograms, and subsequently provide a definitive diagnosis.

Frequency of Operators Used by Participants

The frequency of operator use by participants across levels of expertise and experimental conditions is provided in Table 1. The number of operators used was calculated by selecting 40 protocols at random from a pool of 160 protocols (200 minus 40 used for developing the coding scheme). The coding scheme was then applied to the selected protocols, and a breakdown of the use of operators is presented in Table 1.

Table 1. Frequency of Operator Use by Level of Expertise and Experimental Condition.

	Staff Radiologists		Radiology Residents	
Operators	Authentic Condition N (%)	Augmented Condition N (%)	Authentic Condition N (%)	Augmented Condition N (%)
Data Acquisition	20 (20)	20 (18)	20 (17)	20 (14)
Data Identification	2 (2)	3 (3)	3 (2)	5 (4)
Data Assessment	3 (3)	5 (5)	1 (1)	3 (2)
Data Examination	21 (21)	24 (22)	24 (20)	30 (21)
Data Exploration	17 (17)	23 (21)	35 (29)	34 (24)
Data Comparison	1 (1)	5 (5)	12 (10)	8 (6)
Data Classification	11 (11)	5 (5)	5 (4)	14 (10)
Data Explanation	2 (2)	2 (2)	1 (1)	2 (1)
Hypothesis Generation	19 (19)	15 (14)	16 (13)	19 (13)
Hypothesis Evaluation	4 (4)	1 (1)	0 (0)	3 (2)
Summarization	1 (1)	6 (6)	4 (3)	3 (2)

Overall, residents used more operators than the staff. Both groups used more operators when solving cases presented in the augmented condition. An analysis of the overall frequency of operator use revealed a predominant use of the following operators (listed in order of descending frequency): (1) data examination, (2) data acquisition, (3) data exploration, and (4) hypothesis generation. These four operators account for 76% (88 out of 101 operators and 82 out of 109

operators) of all operators used by the staff in solving cases under both experimental conditions, 79% (107 out of 121 operators) of all operators used by residents solving cases presented under the authentic condition, and 72% (117 out of 143 operators) of all operators used by residents solving cases presented under the augmented condition.

Frequency of Use of Control Processes

The frequencies of control processes is illustrated in Table 2. The frequencies of operator use were calculated from the same selected protocols. Overall, the table shows that staff used slightly more control processes than the residents (47 as compared to 41). However, staff used more control processes when solving cases under the augmented condition than they did under the authentic condition (28 as compared to 19). In contrast, residents used more control processes when solving the cases presented under the authentic condition (23 as compared to 18). Diagnostic planning was the most often used control process regardless of experimental condition, followed by goals. The staff used two-thirds of the goal operators when solving cases under the augmented condition (10 as compared to 5). None of the residents in the sample used goals. The other two control processes used by participants included self-evaluation of diagnostic strategy and experiential memory.

Number of Radiological Findings, Observations, and Diagnoses

The means and standard deviations for the performance measures including the number of radiological findings, observations, and diagnoses are presented in Table 3. The non-statistical comparisons of the means suggests there are no differences between the mean number of radiological findings, observations and diagnoses between the groups across the two experimental conditions. On average, participants identified at least one radiological finding, three radiological observations, and one diagnosis per case. The results do not support the hypothesis that highlighting mammographic cues would facilitate the residents ability to identify findings and observations.

Table 2. Frequency of Control Processes Used by Level of Expertise and Experimental Condition.

	Staff Radiologists		Radiology Residents	
Control Processes	Authentic Condition N (%)	Augmented Condition N (%)	Authentic Condition N (%)	Augmented Condition N (%)
Diagnostic Plans	14 (74)	16 (57)	22 (96)	18 (100)
Goals	5 (26)	10 (36)	0 (0)	0 (0)
Self-Evaluation of Diagnostic Strategy	0 (0)	1 (3.5)	1 (4)	0 (0)
Experiential Memory	0 (0)	1 (3.5)	0 (0)	0 (0)

Table 3. Means and Standard Deviations for the Performance Measures by Level of Expertise Across Experimental Conditions per Case

	Staff Radiologists		Radiology Residents	
Dependent Measures	Authentic Condition	Augmented Condition	Authentic Condition	Augmented Condition
Radiological Findings	1.1 (0.2)	1.1 (0.3)	1.1 (0.2)	1.2 (0.1)
Radiological Observations	2.9 (1.3)	3.2 (1.0)	3.4 (1.3)	3.1 (1.0)
Diagnoses	1.3 (0.3)	1.3 (0.2)	1.2 (0.3)	1.3 (0.2)
Scanning Time (sec)	46.1 (17.6)	47.2 (22.2)	62.8 (19.6)	65.6 (19.6)
Reading Time (sec)	176.7 (50.4)	174.8 (45.5)	199.6 (39.9)	197.4 (33.5)

Scanning Time for Data Acquisition

Scanning time was defined as the amount of time it took a participant to read the clinical history, remove the set of mammograms from the envelope, place them on the viewbox, inspect them with either the naked eye and/or the magnifying glass, and produce his or her first utterance (see Table 3). This operational definition is "loose" in comparison to those specified in eye movement studies, which provide precise operational definitions and use sophisticated equipment required to record scanning time. A repeated measures ANOVA was performed to calculate the level of statistical significance based on level of expertise and by experimental condition on the mean scanning time. Results indicated a significant main effect for expertise, $F(1,18) = 4.89$ $p < .05$, although there was no significant main effect for experimental condition ($F = 0.63$, $p > .05$)

and no interaction ($F = .05$, $p > .05$). As expected, the staff radiologists were significantly faster than residents in scanning the mammograms. The means for scanning time indicate that, on average, the residents took longer (approx. 65 sec) than staff (approx. 47 sec) to scan each breast disease case.

Reading Time for Diagnosis

Reading time was defined as the total amount of time it took a participant to solve a breast disease case, from the initial reading of the clinical history until the last utterance made by the participant while solving the case (see Table 3). A repeated measures ANOVA was performed to calculate the level of statistical significance based on level of expertise and by experimental condition on the mean reading time. Results did not indicate a significant main effect for expertise (F [1,18] = 1.57, $p > .05$) or condition ($F = 0.11$, $p > .05$), and there was no interaction ($F = .0009$, $p > .05$). Overall, the results do not support the hypothesis that staff radiologists would read the cases faster than the radiology residents.

Overall Diagnostic Accuracy

The total number and percentages for overall diagnostic accuracy provided by both groups across experimental conditions was calculated. Overall diagnostic accuracy includes the combination of diagnoses and radiological recommendations. For example, a diagnosis of a carcinoma followed-up by an excisional biopsy would constitute an *accurate* overall diagnosis. In contrast, a diagnosis of a benign lesion followed-up by a biopsy would constitute an *inaccurate* overall diagnosis. Twenty-five percent of the participants (including three staff and two residents) correctly diagnosed and provided the correct subsequent recommendations for the ten breast disease cases.

Again there were minimal differences in the frequencies for overall diagnostic accuracy across groups and experimental conditions. The staff radiologists provided: (1) fewer correct overall diagnoses (41% as compared to 43%), (2) more indeterminate overall diagnoses (6% as compared to 3%), and (3) less inaccurate overall diagnoses (4% as compared to 5%). The small number of correct, indeterminate, and wrong cases was not sufficient to conduct log-linear

analyses across level of expertise and by experimental condition. Therefore, a 2 X 2 Chi-Square analysis was performed on the number of correct and wrong overall accuracy ratings across levels of expertise and experimental conditions (by collapsing indeterminate and wrong errors together). The analysis revealed a non-significant difference in the distribution of the number of cases across levels of expertise and correctness of overall diagnostic accuracy (χ^2 [1, N = 200] = .57, p > .05).

Frequency of Errors

An analysis of the 200 protocols revealed that 34 (17%) errors were committed by staff and residents across the two experimental conditions (see Table 4). The staff made a slightly larger number of errors than residents in the overall diagnostic accuracy (19 as compared to 15), and the residents received slightly more incorrect ratings than the staff (9 as compared to 8). The staff received almost twice the number of indeterminate ratings than the residents (11 as compared to 6).

Table 4. Frequency of Overall Accuracy Errors Committed by Level of Expertise and Experimental Condition.

	Staff Radiologists		Radiology Residents	
Type of Errors	Authentic Condition	Augmented Condition	Authentic Condition	Augmented Condition
Wrong Error	4	4	7	2
Indeterminate Error	6	5	4	2

Of the 34 errors, 17 were coded as wrong and 17 as indeterminate. The staff gave eight of the 17 wrong answers while the residents gave nine. The staff committed an equal number of errors regardless of experimental condition, while residents committed seven errors in the authentic condition and two in the augmented condition.

Of the 17 indeterminate answers, the staff committed eleven and the residents committed six. The staff committed six errors in the augmented condition and five under the augmented condition. The residents, however, committed four errors in the authentic condition and two in the augmented condition.

The staff committed nearly the same number of wrong and indeterminate errors regardless of experimental condition. However, the results suggest that the residents benefited from the highlighting since they committed less errors (both incorrect and indeterminate) when the cases were presented under the augmented condition. For example, they received seven incorrect ratings under the authentic condition (as compared to 2 under the augmented condition), and 4 indeterminate ratings under the authentic condition (as compared to 2 under the augmented condition). These results suggest that the residents were less likely to commit errors when mammographic findings were highlighted. A detailed analysis of the types of errors committed by participants is presented below.

Analysis of Errors Based on Overall Diagnostic Performance.

The frequency and types of errors by level of expertise and experimental condition are presented in Table 5. An in-depth analysis of the 34 errors committed by participants revealed five types of the errors including: (1) a *perceptual detection error* (failure to detect a mammographic finding); (2) a *finding mischaracterization error* (incorrect characterization of a mammographic finding); (3) a *no diagnosis error* (detection, correct identification, and characterization of a mammographic finding but a failure to make a diagnosis); (4) a *wrong diagnosis error* (detection, correct identification, and characterization of a mammographic finding but proposing a wrong diagnosis); and, (5) a *wrong recommendation error* (correct detection and characterization of a mammographic finding, and proposing a diagnosis at some level of abstraction, but proposing an inappropriate subsequent investigation). Overall and in descending order of frequency, the results indicate errors consisted of wrong recommendations (38%), perceptual detection (26%), finding characterization (24%), no diagnosis (6%), and wrong diagnosis (6%). The analyses revealed that regardless of level of expertise and experimental condition, the commission of errors was case-related. Furthermore, the results suggest that the nature of the mammographic features are critical in determining the types of errors committed by radiology professionals.

Table 5. Frequency and Types of Errors by Level of Expertise and Experimental Condition.

	Staff Radiologists		Radiology Residents	
Control Processes	Authentic Condition N (%)	Augmented Condition N (%)	Authentic Condition N (%)	Augmented Condition N (%)
Perceptual Detection Error	3 (30)	1 (11)	5 (45)	0 (0)
Finding Characterization Error	1 (10)	1 (11)	4 (36)	2 (50)
No Diagnosis	2 (20)	0 (0)	0 (0)	0 (0)
Wrong Diagnosis	1 (10)	1 (11)	0 (0)	0 (0)
Wrong Recommendation	3 (30)	6 (67)	2 (18)	2 (50)

Summary

A problem solving model of mammogram interpretation has been presented. Non-statistical comparisons of means revealed no group differences in the number of radiological findings, observations, and number of diagnoses across experimental conditions. Repeated measures ANOVAs revealed that staff radiologists scanned the cases significantly faster than residents with no significant main effect for condition and no interaction, and no differences between groups in reading time across experimental conditions. Analyses revealed that both groups regardless of experimental condition (1) used the same types of operators, control processes, diagnostic plans and goals, (2) committed the same number of errors, and (3) committed case-dependent errors. An additional analysis failed to reveal a significant correlation between the number of total mammograms diagnosed in the past and the number of correctly diagnosed cases in this study. Analyses revealed that mammography interpretation was characterized by a predominant use of data-driven or mixed strategies depending on case typicality and clinical experience.

The Cognitive Basis for the Design of the RadTutor

The results of the empirical work have led to the conceptual framework for the development of the RadTutor. The rationale behind the RadTutor design is based on the assumption that a learner's cognitive processes can be modeled, traced, and corrected in the context of problem-solving (Lajoie, 1993; in press; Lillehaug & Lajoie, 1998). The RadTutor incorporates the results of this

study including the model of problem solving in mammogram interpretation, the problem solving strategies used by staff radiologists and radiology residents, and the typical case-related errors. Furthermore, the framework is also based on a critical assessment of the nature of radiology residency training programs, a review and critique of existing computer-based radiology training environments, an analysis of authentic radiology resident teaching rounds, and instructional principles for the design of the prototype (for an extensive review refer to Azevedo et al., 1997; Azevedo & Lajoie, 1998). However, this section will focus on how the results of this study have been utilized to design the RadTutor (see Figure 1).

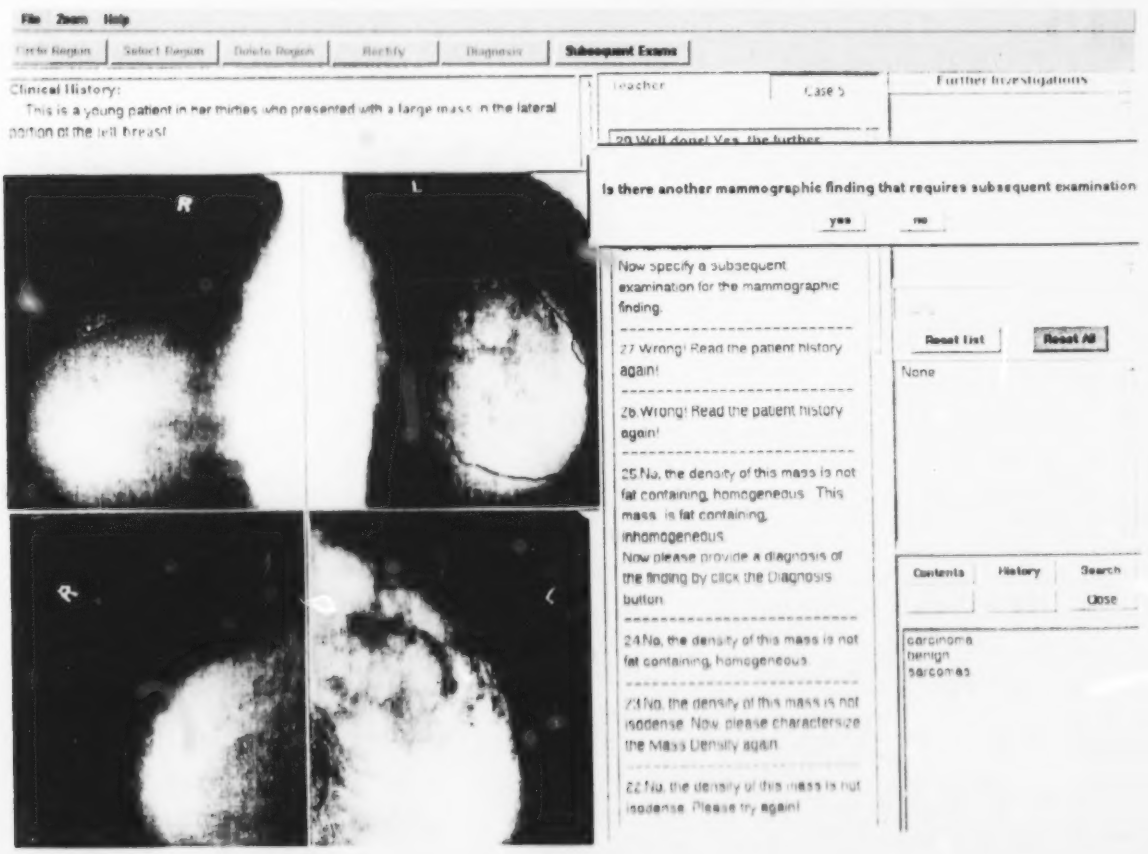
File Zoom Help
Subsequent Exams
Clinical History:
Case 5
Further Investigations
Is there another mammographic finding that requires subsequent examination
yes
no
Now specify a subsequent examination for the mammographic finding.
27 Wrong! Read the patient history again!
26 Wrong! Read the patient history again!
25 No, the density of this mass is not fat containing, homogeneous. This mass is fat containing, inhomogeneous.
Now please provide a diagnosis of the finding by click the Diagnosis button
24 No, the density of this mass is not fat containing, homogeneous.
23 No, the density of this mass is not isodense. Now, please characterize the Mass Density again.
22 No, the density of this mass is not isodense. Please try again!
Reset List
Reset All
None
Contents
History
Search
Close
carcinoma
benign
sarcomas

Figure 1. The RadTutor interface

BEST COPY AVAILABLE

Implications for the Design of the RadTutor

The *content analyses* of the areas of breast disease and mammography have been used to construct the domain knowledge module of the prototype as a series of production rules. The *cognitive task analyses* based on extensive interviews with the domain expert (during case construction) were used to develop the overall instructional sequencing for each case and to build the system's expert module.

The seven-step *problem solving model* characterizing mammogram interpretation served as the overall instructional sequencing for the system. In addition, the system is capable of determining if the user is employing a *data-driven* and/or a *mixed problem solving strategy*. The system monitors the evolution of the user's problem solving behavior (during the resolution of a case) and predicts if he/she is engaged in one or the other problem solving strategies. This aspect of the prototype is extremely critical in identifying errors and providing the appropriate level of scaffolding.

Problem solving operators were used by both radiologists and residents primarily in the design of the different levels of instructional scaffolding and the interface. For example, data acquisition, examination, and exploration meant that the interface should be built to display the case history (data acquisition) and set of mammograms (data acquisition), and to allow the user to manipulate the images for better feature characterization and comparison (data exploration). The system also provides extensive instructional scaffolding during the hypothesis generation phase to ensure that the user has proposed the appropriate level of hypothesis (e.g., malignant versus infiltrating ductal carcinoma).

The verbal protocol analyses indicated that diagnostic planning (i.e., propose further medical examinations) was the most frequent *control process* used by the subjects. As such the interface was built so as to allow the user to list more than one medical examination. This aspect of the prototype is associated with an extensive discussion of the benefits associated with each subsequent examination.

The five types of errors revealed in the *error analyses* are presently being formalized as production rules and integrated in the expert module. Furthermore, each error type is also associated with a specific instructional scaffolding strategy. For example, a finding mischaracterization error is associated with an instructional strategy that focuses the user's attention on the part of a mammographic finding which was mischaracterized (e.g., the border of a mass). The process of identifying error commission is facilitated by the fact that the analyses indicated errors to be case-dependent. For example, cases with atypical mammographic manifestations are highly likely to produce a finding mischaracterization error.

The RadTutor was designed based on the analyses presented in this paper as well as on a critical assessment of radiology residency training programs, a review of existing computer-based radiology training environments, analyses of authentic radiology resident teaching rounds, and instructional principles derived from the cognitive psychology and instructional psychology literature. Furthermore, this theoretically-driven and empirically-derived approach can serve as a generic framework from which to build computer-based learning environments in other educational and professional domains.

Scientific and Educational Importance of the Study

Cognitive research in the area of diagnostic radiology is still in its infancy compared to the corpus of research in other visual (e.g., chess) and medical (e.g., cardiology) domains. As a result, this study is of scientific importance because it provides: (1) an initial characterization of the diagnostic processes involved in mammogram interpretation between radiology professionals with different levels of expertise, (2) a specification of the data-driven (i.e., bottom-up) and mixed (i.e., combination of both bottom-up and top-down) strategies involved in visual diagnostic reasoning, (3) an understanding of how certain factors (e.g., case typicality, case presentation) influence the diagnostic behavior (e.g., problem solving strategies, types of errors committed), (4) an understanding of several performance measures, (5) information about the role of perceptual and problem solving processes during mammogram interpretation, (6) rich empirical data that has the potential to improve training methods, and (7) the empirical basis for the design of the RadTutor.

In conclusion, it is proposed that further research and development endeavors in educational (e.g., math) and professional domains (e.g., auditing) have the potential to benefit from a convergence of theoretical perspectives (e.g., IP, situated cognition) and methodological approaches (e.g., verbal protocol and discourse analyses). Such an approach should be pursued in the area of mammogram interpretation whereby the goal would be to build a comprehensive model of the perceptual and cognitive processes underlying mammogram interpretation. Au such, several researchers are presently conducting research in the area of expert-novice differences in mammography by using: (1) reaction times to assess detection abilities, (2) sorting tasks to assess underlying knowledge structures, (3) longitudinal studies to assess the quantitative and qualitative changes of emerging knowledge structures and problem solving strategies during the course of an individual's entire medical career, and (4) conversational and gestural analyses of teaching rounds focusing on how staff radiologists frame tutoring sessions, ask questions, aid students during problem solving, and react to student errors (verbally and non-verbally). In sum, this study and present research efforts are aimed at furthering our understanding of the interaction between perceptual and cognitive factors underlying mammogram interpretation and improving future radiological training.

Acknowledgments

This paper was prepared for publication while the author was working at the Department of Psychology of Carnegie Mellon University as a postdoctoral fellow funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). This work was also supported by doctoral fellowships from the Fonds pour la Formation de Chercheurs et l'Aide a la Recherche (FCAR), SSHRC, Quebéc government's Ministry of Industry, Science, Commerce, and Technology (MISCT), SAFARI project, and a McGill University grant. I would like to thank my doctoral advisor - Susanne P. Lajoie, and the members of my doctoral committee - Robert Bracewell, Alan Breuleux, David M. Fleiszer, and Alenoush Saroyan. I would also like to thank Sonia Faremo, Jim Greer, Helene Fournier, Patrice Bret, Monique Desaulniers, and the radiologists and residents for their participation in this study.

References

Anderson, J.R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Anderson, J.R., Corbett, A.T., Koedinger, K.R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences, 4* (2), 167-207.

Azevedo, R., & Lajoie, S.P. (1998). The cognitive basis for the design of a mammography interpretation tutor. *Journal of Artificial Intelligence in Education, 9*(1/2), 32-44.

Azevedo, R., Lajoie, S.P., Desaulniers, M., & Fleiszer, D.M. (1996, August). *Tutoring complex visual concepts in radiology*. Paper presented at the 26th International Congress of Psychology, Montreal, Canada.

Azevedo, R., Lajoie, S.P., Desaulniers, M., Fleiszer, D.M., & Bret, P.M. (1997). RadTutor: The theoretical and empirical basis for the design of a mammography interpretation tutor. In B. du Boulay & R. Mizoguchi (Eds.), *Frontiers in Artificial Intelligence and Application* (pp. 386-393). Amsterdam: IOS Press.

Bracewell, R.J., & Breuleux, A. (1993). Substance and romance in analyzing think-aloud protocols. In P. Smagorinsky (Ed.), *Speaking about writing: Reflections on Research Methodology* (pp. 55-88). Newbury Park, CA: Sage.

Breuleux, A. (1991). The analysis of writers' think-aloud protocols: Developing a principled coding scheme for ill-structured tasks. In G. Denhiere & J.P. Rossi (Eds.), *Text and text processing* (pp. 333-362). North-Holland: Elsevier.

Chi, M.T.H., Glaser, R., & Farr, M.J. (1988). *The nature of expertise*. Hillsdale, NJ: Erlbaum.

Ericsson, K.A. (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Erlbaum.

Ericsson, K.A., & Simon, H.A. (1993). *Protocol analysis: Verbal reports as data (2nd ed.)*. Cambridge, MA: MIT Press.

Faremo, S. (1997). *Novice diagnostic reasoning in a visual medical domain: Implications for the design of a computer-based instructional system for undergraduate medical education*. Unpublished master's thesis, Concordia University, Montreal, Quebec, Canada.

Feltovich, P.J., Ford, K.M., & Hoffman, R.R. (1997). *Expertise in context*. Menlo Park, CA: AAAI Press.

Frederiksen, C.H. (1975). Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology, 7*, 371-458.

Gaudette, L.A., Silberberger, C., Altmayer, C.A., & Gao, R.N. (1996). Trends in breast cancer incidence and mortality. *Health Reports, 8*(2), 29-37.

Goergen, S.K., Evans, J., Cohen, G.P.B., & MacMillan, J.H. (1997). Characteristics of breast carcinomas missed by screening radiologists. *Radiology, 204*, 131-135.

Hassebrock, F., & Prietula, M.J. (1992). A protocol-based coding scheme for the analysis of medical reasoning. *International Journal of Man-Machine Studies, 37*(6), 613-52.

Lajoie, S.P., & Lesgold, A. (1992). Apprenticeship training in the workplace: A computer-coached practice environment as a new form of apprenticeship. In M. Farr & J. Psotka (Eds.), *Intelligent instruction by computer: Theory and practice* (pp. 15-36). New York: Taylor & Francis.

Lesgold, A., Feltovich, P.J., Glaser, R., & Wang, Y. (1981). *The acquisition of perceptual diagnostic skill in radiology* (Tech. Rep. No. PDS-1). Pittsburgh: University of Pittsburgh, Learning Research and Development Center.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. Chi, R. Glaser & M. J. Farr (Eds.), *The nature of expertise* (pp. 311-342). Hillsdale, NJ: Erlbaum.

Newell, A., & Simon, H.A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Patel, V.L., Arocha, J.F., & Kaufman, D.R. (1994). Diagnostic reasoning and medical expertise. In D. Medin (Ed.), *The psychology of learning and motivation (Vol. 31)*(pp. 187-252). San Diego, CA: Academic Press.

Patel, V.L., & Groen, G.J. (1986). Knowledge-based solution strategies in medical reasoning. *Cognitive Science, 10*, 91-116.

Posner, M. (1989). *Foundations of cognitive science*. Cambridge, MA: MIT Press.

Winograd, T. (1983). *Language as a cognitive process*. Reading, MA: Addison-Wesley.

Rogers, E. (1992). *Visual interaction: A link between perception and problem solving*. Unpublished dissertation, Georgia Institute of Technology, Atlanta, GA.

Simon, H.A. (1979). *Models of thought*. New Haven, CT: Yale University Press.

Sternberg, R.J., & Smith, E.E. (1988). *The psychology of human thought*. Cambridge, MA: Cambridge University Press.